心理学报 2020, Vol. 52, No. 10, 1199−1211
Acta Psychologica Sinica

https://dx.doi.org/10.3724/SP.J.1041.2020.01199

θ 频段(4~8 Hz)的活动反映了汉语口语产生中音节信息的加工*

蒋宇宸 蔡 笑 张清芳
(中国人民大学心理学系，北京 100872)

摘 要 大脑的神经振荡往往反映了人类的各种认知活动。语言理解的研究发现大脑 θ 频段的活动与音节的加工密切相关, 但目前尚未有研究者探究语言产生过程中大脑特定频段活动与音节加工的联系。我们通过 EEG 时频分析技术, 采用掩蔽启动范式, 考察了 23 名健康被试在图片命名过程中 θ 频段的活动与音节加工过程的关系。行为结果发现, 当启动词和目标图名称之间存在音节相关时, 个体的命名反应时快于音节无关的条件, 而音素相关条件的命名反应时慢于音素无关条件。时频结果发现, 在刺激出现后的 270~460 ms, 音节相关条件下个体 θ 频段神经振荡的能量显著低于音节无关条件, 音素相关与音素无关之间无显著差异。综上, 我们认为汉语口语产生中大脑 θ 频段的活动反映了对音节的加工, 从神经振荡指标上为音节是汉语口语词汇产生中音韵编码的单元提供了证据。

关键词 口语产生; 音节; θ 频段; 时频分析
分类号 B845

1 引言

口语产生是指将思想通过发音器官进行表达的过程(张清芳, 杨玉芳, 2003b), 通常经历概念化(conceptualization)、言语组织(formulation)以及发音运动(articulation)三个阶段。首先, 讲话者需要在概念层面明确自己想要表达的内容, 其次, 对特定的概念信息进行组织并建立相应的发音运动程序, 包括词条选择(lexical selection), 词素音位编码(morphophonological encoding), 音韵编码(phonological encoding)以及语音编码(phonetic encoding)。最后, 通过声带运动将发音目标以声音的形式输出(Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997)。音韵编码阶段的加工单元是口语产生研究的争论焦点之一。

1.1 口语产生中音韵编码单元的跨语言差异

研究表明印欧语系中音素是音韵编码阶段首先被提取的加工单元。Dell (1986)发现在字母语言中发生的语误现象主要为音素遗漏或音素替换。采用内隐启动范式, 研究者发现连续命名首音素相同的目标词(例如 cible-cintre-cerf), 其反应时快于首音素不同的条件, 出现了首音素促进效应(Alario, Perre, Castel, & Ziegler, 2007; Damian & Bowers, 2003; Jacobs & Dell, 2014; Meyer, 1991)。采用形容词−名词命名任务, Damian 与 Dumay (2007)发现被试命名首音素相同的词对(例如 green goat)快于首音素不同的词对(例如 green rug)。另外, 利用图词干扰范式(Damian & Martin, 1999)与掩蔽启动范式(Forster & Davis, 1991; Schiller, 2008), 研究者都发现了首音素促进效应。

然而, 研究表明汉语口语产生中的音韵编码单元是音节而非音素。汉语中的语误现象主要为音节交换错误(Chen, 2000)。Chen, Chen 和 Dell (2002)采用内隐启动范式, 发现音节同质条件比音节异质条件的命名反应时更短, 但音素同质条件与音素异质条件之间无显著差异。利用掩蔽启动范式(Chen,

收稿日期: 2019-10-23
* 北京市社科基金重点项目(16YYA006), 中国人民大学预研委托(团队基金)人才培育类项目(18XNLG28)资助。
通信作者: 张清芳, E-mail: qingfang.zhang@ruc.edu.cn

O'Seaghdha, & Chen, 2016)和图词干扰范式(张清芳, 杨玉芳, 2005), 研究者发现了相同模式的结果。岳源和张清芳(2015)结合图词干扰范式与不同的实验任务(即时命名, 延时命名, 延时命名与发音抑制相结合), 发现音节促进效应发生在音韵编码阶段, 并且该促进作用是一个稳定可靠的效应(Cohen d = 0.85 > 0.8) (Cohen, 1988)。上述研究结果一致地表明音节是汉语口语词汇产生中音韵编码阶段最先被提取的单元。

根据跨语言研究的不同结果, O'Seaghdha, Chen 和 Chen (2010)提出了有关音韵编码的合适单元假说(proximate units principle)。合适编码单元(proximate units)指的是激活词汇词素之后首先被加工的音韵编码单元, 最先选择的单元存在语言上的差异：印欧语系如英语或荷兰语中最先选择的单元是音素, 而在汉语中则为音节。在印欧语系中, 讲话者在选择音素后, 结合节律信息进行音节化过程, 从心理音节表中提取音节准备发音运动程序。汉语口语产生中讲话者在选择音节后进一步分解为音素或音段信息(音韵编码), 准备发音运动程序(语音编码), 最后进行发音, 输出口语产生的结果(发音运动) (同见 Roelofs, 2015)。

音韵编码单元的跨语言差异与不同语言各自的特点密切相关。在汉语中, 音节是受具体语言的语义和结构制约的最小自然发音单位, 对口语产生具有重要意义(张清芳, 2005)。一方面, 汉语的音节数量远远低于其他字母语言(张清芳, 杨玉芳, 2005), 另一方面, 汉语的音节边界相对清晰, 不存在字母语言中重新音节化的现象。因此, 对汉语而言, 更加经济高效的加工方式是将词条的音节信息储存在长时记忆中, 并在音韵编码的早期进行直接提取。而在印欧语系中, 音节的数量巨大, 在发音过程中存在大量重新音节化的现象(Levelt et al., 1999), 因此, 讲话者在音韵编码阶段首先提取的加工单元是音素。

研究者采用事件相关电位(event-related potential, ERP)技术对该问题进行了研究。Qu, Damian 和 Kazanina (2012)采用首音素重复范式(如首音素重复的"黄盒子"和首音素不重复的"绿盒子"), 在图画呈现后的200~300 ms之间发现了首音素重复效应, 表明音素信息在音韵编码阶段也会被激活(同见 Yu, Mo, & Mo, 2014)。利用图图干扰范式和延迟图画命名任务, Wang, Wong, Wang 与 Chen (2017)发现音节效应发生在目标图呈现后的200~400 ms (音韵编码)以及 400~600 ms (语音编码); 采用掩蔽启动范式, Zhang 和 Damian (2019)发现在目标图呈现后的300~400 ms (音韵编码), 音节相关条件诱发了更小的 ERP 波幅。虽然 ERP 能够区分不同实验效应的时间进程, 但上述研究都忽视了口语产生过程中的神经振荡活动(neural oscillations)。传统的 ERP 分析是对相同实验条件下的神经信号进行叠加平均(Rugg & Coles, 1995)。然而, 脑电信号在经过多次叠加后, 会减弱甚至消除非相位锁定的神经振荡活动(Bidelman, 2015)。本研究拟采用脑电时频分析探索汉语母语者音韵编码加工过程中神经振荡的特点, 特别关注的是对应于音节和音素效应的神经振荡。

1.2 θ 频段(4~8 Hz)神经振荡与音节加工之间的关系

神经振荡被认为是大脑神经元的节律性反应, 包括了 delta (δ, < 4 Hz), theta (θ, 4~8 Hz), alpha (α, 8~13 Hz), beta (β, 13~30 Hz), gamma (γ, > 30 Hz)等频段(Ward, 2003)。研究表明神经振荡与个体的注意、记忆、决策等认知过程有着密切的联系(Fell & Axmacher, 2011; Jensen, Kaiser, & Lachaux, 2007; Klimesch, 2012; Siegel, Donner, & Engel, 2012)。人类的语言活动同样会引发特定频段的神经振荡(Giraud & Poeppel, 2012; Lewis, Wang, & Bastiaansen, 2015)。例如, 当被试加工存在句法错误的材料时, 大脑 β 频段的神经振荡能量会显著降低(Bastiaansen, Magyari, & Hagoort, 2009)。

针对语音信息加工的认知神经机制, 不对称时间采样理论(asymmetric sampling in time, AST)提出大脑的听觉腹侧通路(颞上回 STG—颞上沟 STS—颞下沟 ITS)将按照刺激的声学属性进一步分化成两条平行的加工通路, 分别负责提取音节和音素水平的信息 (Poeppel, 2003)。一段连续语流中所包含的音节信息在时程上的变化速率相对音素而言是比较缓慢的, 而大脑低频的 θ 活动则表征了对音节的加工(Doelling, Arnal, Ghitza, & Poeppel, 2014; Howard & Poeppel, 2012; Luo & Poppel, 2007; Peelle, Gross, & Davis, 2013)。虽然大脑神经元的放电频率和语音信号自身的音节变化频率不是简单的直接对应关系, 但这仍然提示我们 θ 频段的神经振荡活动与音节的加工紧密相关。

第一, 大脑 θ 活动的相位信息表征了对音节的追踪(Ghinst et al., 2016; Gross et al., 2013; Molinaro, Lizarazu, Lallier, Bourguignon, & Carreiras, 2016)。

在 Pefkou, Arnal, Fontolan 和 Giraud (2017)的实验中，研究者首先对人们自然状态下产生的实验材料(句子)音频进行包络分析，得到该声音信号中所包含的音节总数，再将其除以自身的总持续时间，计算出这段音频的音节速率，最后通过对材料进行时程上的压缩，以获得不同音节速率的实验刺激。进入正式实验后，主试在记录被试脑电信号的同时，向被试播放不同音节速率的句子。回归分析结果发现，个体的 θ 相位一致性随着音节速率的升高而降低，也就是说，当句子的播放速度越快，被试越难追踪其中的音节信息时，大脑 θ 活动的相位一致性就越差。此外，研究发现当被试接受连续的音节序列(syllable sequences)刺激时，θ 神经振荡的相位一致性会提高(Power, Mead, Barnes, & Goswami, 2012)。

第二，大脑 θ 活动的能量信息反映了对音节的识别。在一项跨语言的研究中，Peña 和 Melloni (2012)采用日语，西班牙语，意大利语三种不同语言的口语句子作为实验材料考察西班牙语母语者和意大利语母语者语言理解的动态过程。结果发现不管是哪种母语类型的被试，当听到正序播放的句子时，θ 频段神经振荡的能量显著高于逆序播放句子的条件。并且，当被试听非本国语的材料时，对比正序播放和逆序播放两种实验条件，θ 频段的能量活动表现出加工本国语言材料时相同的模式。正序播放和逆序播放的材料虽然保持了基本声学属性的高度一致，但倒放的材料会造成严重的语音扭曲(phonological distortions)，导致个体无法理解材料的意义(Binder et al., 2000; Gross et al., 2013; Saur et al., 2010)。Peña 和 Melloni (2012)指出，实验材料在时程上的反向破坏了各个单词原有的语音结构，造成个体难以切分单词的音节，从而使个体 θ 频段的神经振荡活动变弱。在一项关于汉语的研究中，当研究者向被试播放音节刺激时，正放的条件相比于倒放的条件同样诱发了更强的 θ 频段能量活动(Ding, Melloni, Zhang, Tian, & Poeppel, 2015)。

综上，行为和 ERP 研究均表明音节在汉语口语产生过程中扮演了重要角色，语言知觉或理解的研究表明 θ 频段的神经活动与音节加工密切相关。目前尚未有研究考察汉语口语产生中的音节启动效应是否与特定频段的神经振荡活动相关，而脑电时频分析有助于我们更深入地了解汉语口语产生过程中对于音节的加工机制。本研究中我们采用掩蔽启动范式，操纵启动词和目标图名称之间的语音相关关系(音节相关−音节无关，音素相关−音素无关)，要求被试完成图画命名任务，同时记录其脑电信号，最后对 EEG 数据进行时频分析，考察音节效应和音素效应发生时的神经振荡。在掩蔽启动范式中，启动词的呈现时间非常短(50 ms 左右)，被试对启动刺激的加工通常是阈下的，该范式排除了个体命名策略等无关因素对实验结果的影响(Chen et al., 2016; You, Zhang, & Verdonschot, 2012)。由于 θ 频段和音节的加工存在密切联系(Ghinst et al., 2016; Gross et al., 2013; Power et al., 2012; Peña & Melloni, 2012)，我们预期会发现显著的音节启动效应，但不会发现显著的音素启动效应，相应地，θ 频段的能量仅在音节相关和无关条件中会存在显著差异。

2 方法

2.1 被试

23 名大学生和研究生(11 名男生，平均年龄 22 岁)。所有被试均为右利手，无任何精神疾病病史，母语为汉语，讲标准普通话，视力或矫正视力正常。在参加实验之前被试阅读知情同意书并签字，实验之后获得一定报酬。

2.2 材料

64 幅由黑白线条组成的图片，选自张清芳和杨玉芳(2003a)建立的汉语图片库，其中用于正式实验的图片 60 幅，练习试次图片 4 幅。每张图片的名称为双音节名词，第一个字分别与 4 种实验条件匹配。例如，图片名称为“鼻子” (/bi2zi5/)，音节相关条件的启动字为“彼” (/bi3/)，与图片名称音节完全相同但声调不同；音素相关条件的启动字为“柏” (/bai3/)，与图片名称的首音素相同，声调不同。音节相关与音素相关的条件下启动词的词频(Cai & Brysbaert, 2010)没有显著差异(t = 0.027, p = 0.978)。音节相关条件与音素相关条件的启动字在随机打乱后重新与图片进行匹配形成音节无关条件与音素无关条件。检查后，两种无关条件下的启动字与目标图名称之间不存在语音相关。4 种实验条件下启动词和图片名称之间不存在语义和正字法相关。

2.3 设计

研究采用 2×2×2 被试内设计，自变量包括相关类型(音节，音素)、相关条件(相关，无关)和重复次数(第一次，第二次)。每幅图片与 4 种不同类型启动字匹配，因此每组测试中包括了 240 个试次。每

组测试重复两次，因此每个被试共完成 480 个试次。每个被试在每组测试中的试次呈现顺序都是不同的，通过伪随机的方式呈现，保证相同的图片之间至少间隔 5 个试次，图片名称首音素相同的试次不会连续出现。

2.4 实验仪器

E-prime 2.0 编写实验程序，PST SRBOX 反应盒，麦克风与计算机。实验图片均通过计算机呈现在屏幕中央，被试的反应通过反应盒连接的麦克风记录。实验材料的呈现和被试的反应时由电脑控制与收集。主试记录被试是否进行正确反应。采用国际通用 10-20 系统的 64 导脑电帽，NeuroScan 系统记录被试的脑电信号。

2.5 程序

实验分为 3 个阶段，学习阶段、测试阶段和正式实验阶段。在学习阶段，屏幕中央会依次呈现每幅图片及其对应的名称 2 s。主试告知被试接下来正式实验中会出现这些图片并要求被试记住图片的内容及其对应的名称。在测试阶段，呈现图片要求被试说出图片名称，当被试对所有图片都能正确命名时，方可进入正式实验。所有材料都是日常生活中常见的且命名一致性较高的图片，所有被试均能顺利完成对图片名称的学习。

正式实验中每个试次的流程如下(见图 1)：首先呈现 500 ms 的注视点(“+”)，然后是 500 ms 的前掩蔽(@@)，接着会呈现 49 ms 的启动词，随后是 20 ms 的后掩蔽(@@)，掩蔽消失后，屏幕中央会呈现目标图片。被试需要在 2000 ms 内又快又准地对图片进行命名，做出反应的同时图片会立刻消失，间隔 1000 ms 后开始下一次测试。正式实验之前有 4 次练习使被试熟悉实验任务。所有的图片刺激均标准化为统一大小，启动字为 28 号宋体，掩蔽刺激(@@)为 36 号宋体。完成对所有试次的命名需要 100 分钟。

2.6 脑电记录与分析

在线记录以左侧乳突作为参考电极，离线分析

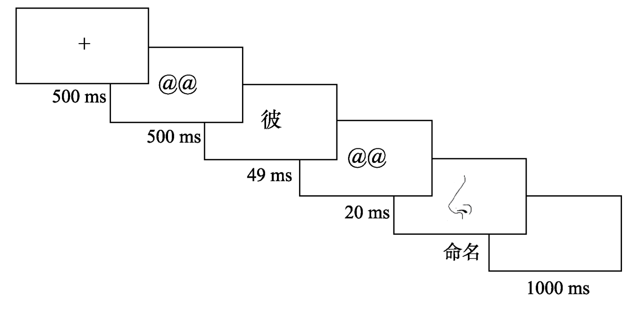


图 1 掩蔽启动范式实验流程

时重参考为双侧乳突。同时记录双眼的水平眼电(HEOG)和左眼的垂直眼电(VEOG)。电极与头皮之间的阻抗均小于 5 kΩ，滤波带通为 0.05~70 Hz，信号采样率为 500 Hz。

预处理采用 EEGLAB 工具包进行(Delorme & Makeig, 2004)。首先，对于信号采集时与头部接触不良或已损坏的电极(数量未超过总电极数的 5%)，采用 EEGLAB 中自带函数进行坏导替换，通过“球面插值算法” (spherical interpolation)，利用被替换导联周围电极的数据，对信号重新估计后进行替换(Perrin, Pernier, Bertrand, & Echallier, 1989; Pivik et al., 1993)。第二，对数据进行 0.1~30 Hz 的滤波以及独立成分分析(independent component analysis, ICA)。ICA 基于盲源信号分离技术，基本思路是将多通道观察的信号按照统计独立的原则分解为若干成分(ICs)。ICA 对脑电信号中由于眨眼，肌肉运动引发的伪迹有较高的识别度(Makeig, Bell, Jung, & Sejnowski, 1995)。眼电伪迹的判断标准为：成分排序靠前，成分的头部活动集中在前额区，能量活动随频率升高缓慢衰减，单试次能量大。肌电伪迹的判断标准为：成分的头部活动分散在外侧的局部地区，能量活动随频率升高而升高，单试次能量较大。按上述标准对每个成分依次进行检查，并结合 EEGLAB 识别伪迹成分的插件 ADJUST，对伪迹进行排除。第三，按图片出现前 1000 ms 以及出现后 1500 ms 对脑电信号进行分段，在排除波幅超过 ± 100 μV 的试次以及肉眼确认后，将数据保存。

利用 Fieldtrip 工具包对数据进行时频分析，采用小波变换的方法(Oostenveld, Fries, Maris, & Schoffelen, 2011)。单试次总时长为 2500 ms，时间分辨率为 10 ms；分析频段范围为 3~30 Hz，频率分辨率为 1 Hz。小波周期以线性方式递增，最低频率处周期为 3，最高频率处周期为 8 (同样的标准见 Li, Shao, Xia, & Xu, 2019)。随后将不同周期的正弦小波与脑电时域信号进行卷积，进而获得不同频率范围内各个时间点脑电信号的神经振荡能量值(Goupillaud, Grossmann, & Morlet, 1984)。分析过程中首先对单个试次的活动能量进行估计，再完成多试次之间的平均。以刺激出现前 300~100 ms 作为基线，采用分贝量尺对能量活动进行校正：dB = 10 × log10 (基线后能量/基线平均能量)。简明起见，事件相关频谱扰动(event-related spectral perturbation, ERSP)呈现的频率范围为 4~20 Hz。

根据电极点分布的空间位置，研究选取了 6 个

兴趣区(regions of interest, ROI)并对每个兴趣区内的神经振荡能量进行平均，分别为：左前区(F3, FC3, FC5)，中前区(Fz, FCz, Cz)，右前区(F4, FC4, FC6)，左后区(P5, P3, PO3)，中后区(CPz, Pz, POz)与右后区(P6, P4, PO4)。采用 2(相关类型：音节，音素) × 2(相关条件：相关，无关) × 2(重复：第一次，第二次) × 6(兴趣区：左前，中前，右前，左后，中后，右后)重复测量方差分析，统计结果非球形性时利用 Greenhouse-Geisser 法对 p 值进行校正。

在比较音节相关和无关条件，音素相关和无关条件的差异是否显著时，我们采用了基于簇的置换检验(cluster-based permutation test)对数据进行统计分析，该方法能够有效地对多重比较下的 p 值进行校正(Maris & Oostenveld, 2007)。进行置换检验的时间窗口为刺激出现后的 600 ms，步长为 10 ms，共 6 个兴趣区，对感兴趣的每两个实验条件之间的数据(时间×频率×电极)进行重复测量 t 检验，p 值小于 0.05 且在时间和空间位置上邻近的数据点将被合并为同一个簇。随后，计算每个簇内 t 值的和以确定簇水平(cluster level)的统计信息，通过蒙特卡洛法(Monte Carlo method)进行统计显著性检验，随机抽样的次数为 1000 次。

3　结果

3.1　行为结果

删除两个错误率大于 10%以及两个反应时均值在 3 个标准差之外的项目。删除设备未记录到的数据(2.5%)以及命名错误的数据(1.8%)。删除反应时小于 200 ms 以及大于 1500 ms 的数据，包含偏离平均值 3 个标准差之外的数据在内，占总试次的 2.3%。图 2 所示为不同实验条件变化下命名反应时的均值和 95%置信区间(confidence interval, CI)。

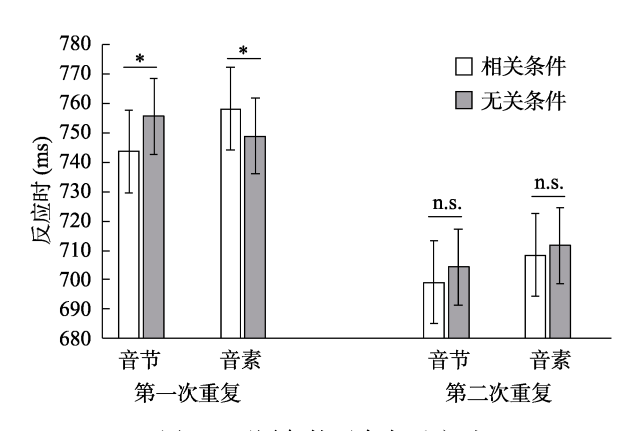


图 2　不同条件下命名反应时
注：误差棒为 95% CI, * $p < 0.05$, n.s.表示无显著差异

对反应时数据进行 2(相关类型：音节，音素) × 2(相关条件：相关，无关) × 2(重复次数：第一次，第二次)重复测量方差分析，被试分析(F_1)与项目分析(F_2)结果表明，相关类型的主效应在被试分析中显著，项目分析中边缘显著：被试在音节条件下的命名反应时更快，$F_1(1, 22) = 7.67$, $p = 0.01$, $\eta_p^2 = 0.26$, $F_2(1, 55) = 3.46$, $p = 0.07$, $\eta_p^2 = 0.06$。重复次数的主效应显著：被试在第二次重复时命名反应时更快，$F_1(1, 22) = 48.24$, $p < 0.001$, $\eta_p^2 = 0.69$, $F_2(1, 55) = 282.59$, $p < 0.001$, $\eta_p^2 = 0.84$。相关类型和相关条件的交互作用显著：$F_1(1, 22) = 5.51$, $p = 0.03$, $\eta_p^2 = 0.20$, $F_2(1, 55) = 7.04$, $p = 0.01$, $\eta_p^2 = 0.11$。简单效应分析表明，被试在音节相关条件下命名反应时显著快于音节无关条件，$F_1(1, 22) = 6.12$, $p = 0.02$, $\eta_p^2 = 0.22$, $F_2(1, 55) = 5.87$, $p = 0.02$, $\eta_p^2 = 0.10$; 音素相关条件与音素无关条件之间命名反应时差异不显著，$F_1(1, 22) = 0.85$, $p = 0.37$, $\eta_p^2 = 0.04$, $F_2(1, 55) = 1.45$, $p = 0.23$, $\eta_p^2 = 0.03$。相关类型，相关条件与重复次数之间的三重交互作用在被试分析中显著，项目分析中边缘显著：$F_1(1, 22) = 7.88$, $p = 0.01$, $\eta_p^2 = 0.26$, $F_2(1, 55) = 3.75$, $p = 0.06$, $\eta_p^2 = 0.06$。第一次重复时，相关类型与相关条件的交互作用显著，$F_1(1, 22) = 13.38$, $p = 0.001$, $\eta_p^2 = 0.38$, $F_2(1, 55) = 8.53$, $p = 0.005$, $\eta_p^2 = 0.13$。具体表现为被试在音节相关条件下命名反应时显著快于音节无关条件，$F_1(1, 22) = 5.88$, $p = 0.02$, $\eta_p^2 = 0.21$, $F_2(1, 55) = 6.01$, $p = 0.02$, $\eta_p^2 = 0.10$; 而音素相关条件的命名反应时显著慢于音素无关条件，$F_1(1, 22) = 5.04$, $p = 0.04$, $\eta_p^2 = 0.19$, $F_2(1, 55) = 4.53$, $p = 0.04$, $\eta_p^2 = 0.08$。第二次重复时，相关类型与相关条件的交互作用不显著，$F_1 < 1$, $F_2 < 1$。其他的主效应和交互作用均不显著。

3.2　时频分析结果

一名被试由于信号伪迹过大未纳入后续分析。删除反应时小于 500 ms (5.1%)，大于 1500 ms (1.0%)以及伪迹较大的试次(5.3%)。我们根据相关研究确定了 6 个连续的时间窗口：0~100 ms, 100~200 ms, 200~300 ms, 300~400 ms, 400~500 ms, 500~600 ms (类似的窗口划分见 Qu et al., 2012; Zhang & Damian, 2019)，每个时间窗口内 θ 能量活动方差分析的结果见表 1。

在 0~100 ms, 100~200 ms 和 200~300 ms 时间窗内，相关类型的主效应显著，音节条件诱发了更大的能量活动，$F(1, 21)_{0\sim100\ ms} = 8.08$, $p = 0.01$, $\eta_p^2 = 0.28$, $F(1, 21)_{100\sim200\ ms} = 11.03$, $p = 0.003$, $\eta_p^2 = 0.35$,

表 1 以相关条件、相关类型、重复次数以及兴趣区为自变量 θ 能量活动在 0~600 ms 时间窗内的方差分析

变异来源	0~100 ms		100~200 ms		200~300 ms		300~400 ms		400~500 ms		500~600 ms	
	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
相关类型(1, 21)	8.08*	0.28	11.03**	0.35	5.18*	0.20	n.s.		n.s.		n.s.	
相关条件(1, 21)	n.s.		n.s.		n.s.		n.s.		n.s.		n.s.	
重复次数(1, 21)	n.s.		n.s		n.s.		n.s.		n.s.		n.s.	
相关类型×相关条件(1, 21)	n.s.		n.s		n.s.		5.13*	0.20	6.41*	0.23	6.00*	0.22
相关类型×相关条件×重复次数(1, 21)	n.s.		n.s.		n.s.		n.s.		n.s.		n.s.	
相关条件×相关类型×重复次数×兴趣区(5, 105)	n.s.		n.s.		n.s.		n.s.		n.s.		n.s.	

注：** $p < 0.01$, * $p < 0.05$, n.s. 不显著。

$F(1, 21)_{200\sim300\ ms} = 5.18$, $p = 0.03$, $\eta_p^2 = 0.20$。在 300~400 ms 时间窗内，相关类型和相关条件的交互作用显著，$F(1, 21) = 5.13$, $p = 0.03$, $\eta_p^2 = 0.20$。音节相关条件的平均能量小于音节无关条件，$F(1, 21) = 4.43$, $p = 0.047$, $\eta_p^2 = 0.17$；音素相关与音素无关条件之间无显著差异，$p = 0.14$。在 400~500 ms 时间窗内，相关类型和相关条件的交互作用显著，$F(1, 21) = 6.41$, $p = 0.02$, $\eta_p^2 = 0.23$。音节相关条件的平均能量小于音节无关条件，$F(1, 21) = 5.81$, $p = 0.03$, $\eta_p^2 = 0.22$；音素相关与音素无关条件之间无显著差异，$p = 0.38$。在 500~600 ms 时间窗口内，相关类型和相关条件交互作用显著，$F(1, 21) = 6.00$, $p = 0.02$, $\eta_p^2 = 0.22$。音节相关条件的平均能量小于音节无关条件(边缘显著)，$F(1, 21) = 3.57$, $p = 0.07$, $\eta_p^2 = 0.15$；音素相关与音素无关条件之间无显著差异，$p = 0.16$。其他频段的能量活动在各个时间窗内的交互作用都不显著。

针对音节效应和音素效应所进行的基于簇的置换检验结果显示，第一次重复时，右前兴趣区(F4, FC4, FC6)音节相关条件比音节无关条件在刺激出现后的 270~460 ms 之间诱发了更低的 θ 频段能量活动($p = 0.01$, 4~8 Hz)；左前兴趣区(F3, FC3, FC5)音素相关条件比音素无关条件在刺激出现后的 340~390 ms 诱发了更高的 θ 频段能量活动(边缘显著, $p = 0.052$, 4~8 Hz)。第二次重复时，各个兴趣区的音节效应，音素效应均不显著。在 3~30 Hz 之间其他的频段(δ, α, β)的能量活动上，感兴趣的实验条件之间差异不显著(见表 2)。图 3a 与 3b 分别呈现了不同重复次数下音节效应与音素效应的 ERSP。

4 讨论

本研究采用掩蔽启动范式考察了汉语口语产生中音节与音素的加工过程与 θ 频段能量活动之间的关系。行为结果发现个体在音节相关条件下对图片的命名快于音节无关条件，而在音素相关条件下命名反应时慢于音素无关条件。脑电时频分析结果表明，在图片出现后的 270~460 ms, θ 频段的能量

表 2 0~600 ms 之间各频段能量活动的音节效应与音素效应

实验效应		频段	0~300 ms	300~600 ms
重复一	音节效应	δ, α, β	n.s.	n.s.
		θ	270~300 ms $p = 0.01$	300~460 ms $p = 0.01$
	音素效应	δ, α, β	n.s.	n.s.
		θ	n.s.	340~390 ms $p = 0.052$
重复二	音节效应	δ, α, β	n.s.	n.s.
		θ	n.s.	n.s.
	音素效应	δ, α, β	n.s.	n.s.
		θ	n.s.	n.s.

0.3 dB
−0.2 dB

注：n.s. 表示差异不显著($p > 0.1$), p 值经簇水平校正

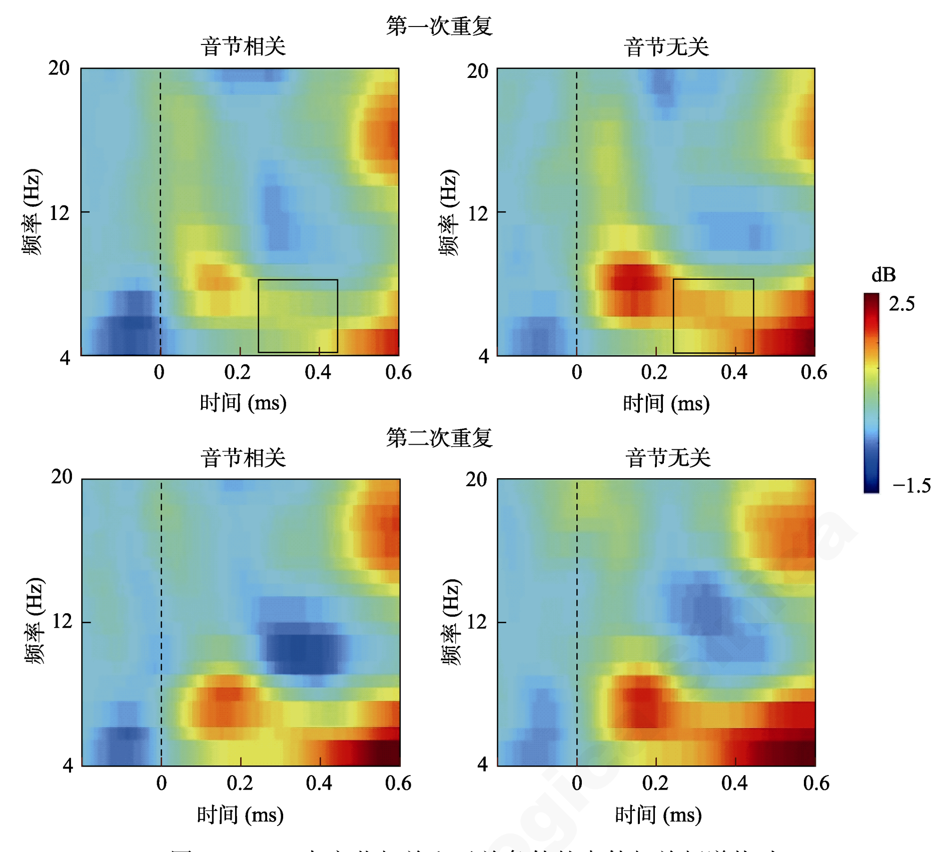


图 3a　FC4 点音节相关和无关条件的事件相关频谱扰动

注：虚线代表图片出现的时间，黑色方框表示条件之间差异显著，$p < 0.05$，经簇水平校正

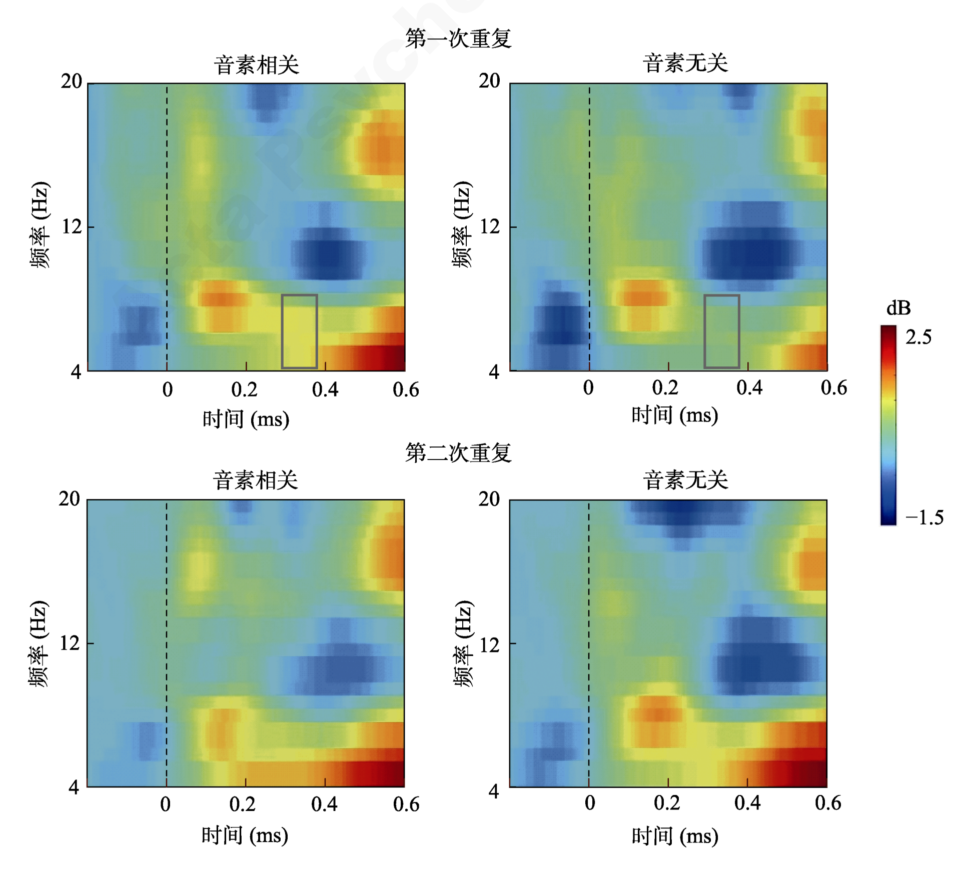


图 3b　FC3 点音素相关和无关条件的事件相关频谱扰动

注：虚线代表图片出现的时间，灰色方框表示条件之间差异边缘显著，$0.05 < p < 0.1$，经簇水平校正

活动在音节相关与音节无关条件之间差异显著，表现为音节相关条件下个体 θ 频段的能量更低，音素相关条件与音素无关条件相比无显著差异。研究结果表明，汉语口语产生过程中，θ 频段的活动反映了个体对音节信息的加工。

在行为层面，与 Zhang 和 Wang (2014)的研究结果一致，由于试次重复带来的练习效应不仅总体降低了第二次命名的反应时，也使得实验关注的效应消失了。重要的是，在第一次命名时，本研究成功重复了以往利用内隐启动范式(Chen et al., 2002)，图词干扰范式(岳源，张清芳，2015)以及掩蔽启动范式(Chen et al., 2016)发现的音节促进效应；同时，我们发现个体在音素相关条件下命名潜伏期更长，表现出音素抑制效应(Chen et al., 2016)。需要注意的是，在以往针对字母语言的研究中，呈现首音素相关的启动刺激会降低个体对目标刺激的命名反应时(Alario et al., 2007; Damian & Bowers, 2003; Damian & Martin, 1999; Forster & Davis, 1991; Schiller, 2008)。这提示我们字母语言和汉语的口语产生过程是不同的，具体而言，字母语言需要在音韵编码阶段通过加工刺激的重音、音素等信息确定其音节结构(Levelt et al., 1999)；而汉语的音节数量少，音节边界清晰，不存在重新音节化的特点，这使得汉语母语者在音韵编码阶段可以直接提取刺激的音节信息(O'Seaghdha et al., 2010)。结合本研究的反应时结果来看，个体表现出的音节促进效应表明汉语口语产生中音节能够作为独立的信息表征单元，在音节相关条件下通过提前加工目标刺激的音节信息降低了命名反应时；而音素抑制效应很有可能是相同首音素的音节协同激活(co-activation)造成的竞争导致的。在音素相关条件下，由于启动刺激与目标刺激的首音素是相同的，部分重合的信息可能导致以该音素开头的音节都得到了激活，致使个体在提取目标词的音节信息时，受到了来自其他音节信息的干扰，延长了命名反应时(后文将结合 θ 频段的神经振荡活动对音素抑制效应的竞争机制展开具体讨论)。类似的，在一项针对汉语的听觉词汇判断任务中，Sereno 和 Lee (2015)同样发现了音素抑制效应。尽管如此，相比于音节促进效应而言，我们认为本研究的音素抑制效应是较为微弱的：个体在第二次测试时没有表现出显著的音节促进效应和音素抑制效应，但相关类型与相关条件的两因素交互作用依旧是显著的，个体在音节相关条件下命名反应时更快，而音素条件之间差异并不显著。这至少说明再次命名时，被试的行为表现更倾向于音节促进效应，而非音素抑制效应(见图 2)。

在神经层面，我们发现 θ 频段的神经振荡活动在相关类型与相关条件之间存在交互作用：音节相关条件相比于音节无关条件，个体 θ 频段的能量更低。更重要的是，音节效应发生在刺激出现后大约 270~460 ms，与以往图片命名 ERP 研究中音节效应的时间窗口一致(Cai, Yin, & Zhang, 2020; Dell'acqua et al., 2010; Wang et al., 2017; Yu et al., 2014; Zhu, Damian, & Zhang, 2015)。根据元分析的结果(Indefrey & Levelt, 2004)，个体通常在看到图片之后的 250 ms 进入音韵编码阶段，此时汉语母语者需要提取目标词的音节信息并为后续加工做准备。θ 频段的能量在音节相关条件下比无关条件下更低，这存在两种可能的解释。第一，音节重复导致相同音节所引起的能量活动下降。GrillSpector, Henson 和 Martin (2006)指出，个体在加工重复的刺激时神经元活动会减弱，而这种去激活反应表现为神经振荡能量的降低。Gruber 和 Müller (2002)在客体识别的任务中发现，在图片呈现后的 220~350 ms，重复的刺激比非重复的刺激诱发了更低的 γ 频段神经振荡能量。采用任务切换范式，研究者发现当被试完成的任务前后保持一致时(重复条件)，相比不一致(非重复条件)，β 与 γ 的能量显著下降(Gruber, Giabbiconi, Trujillobarreto, & Müller, 2006)。刺激的重复加工引起神经振荡能量减弱这一结果在面孔识别的研究中也得到了重复(Engell & Mccarthy, 2014)。Brookes 等(2005)利用功能磁共振(functional magnetic resonance imaging, fMRI)的技术，发现由重复引起的神经振荡能量下降与皮层血氧依赖水平(blood oxygenation level dependent, BOLD)的衰减密切相关。本研究中，音节相关条件下启动词的音节与目标图名称首字的音节完全一致，这种语音上的相似性极有可能引起"重复效应"，造成 θ 频段能量活动的降低。第二，两种条件下认知加工负荷的差异引起了能量的变化。来自工作记忆的研究证据表明，对于认知加工负荷较重的任务，神经元释放的能量相对较高，而认知加工负荷较轻的任务，神经元释放的能量则相对较低(Roux & Uhlhaas, 2014)。在音节相关条件下，由于对目标图名称的音节信息进行了提前加工，当被试看到目标刺激并提取相应的音节进行音韵编码时，认知加工的负荷是相对较低的。相反，音节无关条件下，被试无法通过启动刺激直接提取目标图片的音节信息，认知加

工负荷相对较高。因此，对比音节无关条件, θ 神经振荡活动的能量在音节相关条件下会更弱，这与行为反应时的模式一致：被试在音节相关条件下认知负荷更低，命名更快。

需要指出的是，在言语知觉过程中，研究者比较自然的语音材料和逆序播放的声音刺激时，他们发现了在自然的语音材料知觉过程中引发了更高的 θ 频段能量活动，这与本研究发现的模式不一致。这是由于比较条件的不同引起的：Peña 和 Melloni (2012)以及 Ding 等(2015)的研究发现在言语知觉过程中，当被试听逆序播放的材料时，相比于正序播放的条件, θ 频段的活动能量更低。研究者认为逆序播放破坏了语言本身正常的语音结构，人们不能识别材料中的音节，相应皮层的神经元活动处于抑制状态，其神经振荡的能量相对较低。在正序播放材料的情况下，被试能够顺利地加工材料的音节信息，相应皮层的神经活动处于兴奋状态，其神经振荡的能量相对较高。我们研究中比较的是音节相关与无关，是两类兴奋状态下的比较，而言语知觉研究中比较的是兴奋状态和非兴奋状态，因此产生了不同的模式。

与音节效应相比，在 340~390 ms 的时间窗内，我们发现了音素相关条件与音素无关的能量差异(边缘显著水平)，表现为第一次重复时音素相关条件下 θ 频段能量活动高于音素无关条件，其能量活动模式与音节相关条件下不同。当启动词和目标图名称存在音素相关时，二者享有相同的首音素(启动词：柏/bai3/，目标图：鼻子/bi2//zi5/)，在加工过程中首音素的重叠激活了以该音素开头的所有音节(如：/ba/, /bang/等)，对目标音节的产生造成了竞争，表现为 θ 频段的能量活动在音素相关条件与音素无关条件对比时更强。这与 Chen 等(2016)的发现一致，他们采用与本研究完全相同的实验设计和任务，发现音素相关与无关条件相比出现了微弱的抑制效应。

在汉语口语词汇产生中，音节效应 θ 频段能量变化的时间早于音素效应，与已有研究中相关效应的时间进程一致，从能量变化的角度为“音节是音韵编码的合适单元”提供了证据。例如，Feng，Yue 和 Zhang (2019)采用图画−词汇干扰任务，操纵了干扰词与目标图名称之间的相关关系，发现音节相关效应出现在图画呈现后的 320 ms 左右，音素相关效应在 368 ms 左右，这两类效应均发生于音韵编码阶段。在 Zhang 和 Damian (2019)的研究中，除了发生在 300~400 ms 的音节效应，研究者还探测到了音素效应：在目标图呈现后的 500~600 ms，音素相关条件比音素无关条件诱发了更小的 ERP 波幅。

虽然目前没有研究者利用脑电技术直接对比印欧语母语者音素效应与音节效应的时间进程，但研究者对于音节频率效应(syllable frequency effect)的研究也许能提供部分的证据。音节频率效应是指个体在语言产生过程中对使用频次较高的音节加工速度更快，反应时更短，该效应与音节水平的加工密切相关(Levelt et al., 1999)。利用 ERP 反应锁时(response locked)的分析方法, Burki, Cheneval 和 Laganaro (2015)发现音节频率效应出现在被试出声命名前的 180~150 ms 之间，该时间窗正好对应口语产生的语音编码阶段。Dell’Acqua 等(2010)通过对比意大利母语者口语产生过程中的语义效应和语音效应，发现首音素重合引发的促进效应发生在刺激出现后的 250~400 ms 之间，即音韵编码阶段(Indefrey & Levelt, 2004)。以上的两个研究在脑电时间进程上为字母语言先加工音素后加工音节提供了间接证据。相比而言，我们发现的汉语口语词汇产生中能量变化上的先后关系为音节效应在前，音素效应在后。汉语和印欧语系中音节和音素的提取在时间进程上表现出完全相反的模式。上述对比表明不同的语言，音韵编码中首先提取的单元不同，与“合适单元假说”的跨语言假设一致(O’Seaghdha et al., 2010)。

音节相关条件缩短图画命名时间，引起 θ 频段能量的减弱，而音素相关条件可能会延长图画命名时间，引起 θ 频段能量的增强。本研究和言语知觉的结果都表明无论是语言理解任务还是语言产生任务, θ 频段的能量活动与音节的加工相关。尽管如此，研究中微弱的处于边缘显著水平的音素效应提示 θ 频段的能量活动可能和语音加工相关。第二次重复时在反应时上音节和音素效应都消失，这与已有的研究一致(Chen et al., 2016)，本研究的时频结果与反应时结果一致，表明人类的大脑能够迅速对加工过的刺激作出反应。下一步的研究中可以采用英语作为目标语言考察在行为结果上出现显著的音素效应时，θ 频段的能量活动是否在相关条件和无关条件下存在差异，进行跨语言之间的对比，深入考察 θ 频段的能量活动的认知涵义。

需要注意的是，虽然传统的 ERP 分析方法在时间进程上区分了音节和音素的加工顺序，但 ERP

的幅值在不同研究中表现出了完全相反的模式：与Zhang 和 Damian (2019)的结果不同, Dell'Acqua 等(2010)发现语音相关条件比语音无关条件诱发了被试更大的 ERP 波幅。与语义违反的"N400 效应"存在矛盾的实验结果类似(van Petten & Luka, 2012), ERP 波形方向和波幅大小在实验间的不一致使研究者很难进一步解释语言加工的脑机制。即使 ERP 的结果表现出较高的跨实验一致性, 由于差异波并不与特定的认知加工过程对应, 我们仍然无法将条件之间不同的 ERP 模式完全归因于口语产生过程中加工机制的不同。例如, 在 Wang 等(2017)的研究中, 音节相关条件在刺激属性上重叠度相对较大(剪刀/**jian**3dao1/—键盘/**jian**4pan2/), 音素相关条件在刺激属性上重叠度相对较小(西瓜/**xi1**gua1/—信封/**xin**4feng1/), 被试进行图片命名时, 完全有可能在阈下水平探测到了不同实验条件之间重叠度的差异, 因此音节相关条件与音节无关条件对比时差异更大, 而音素相关条件与音素无关条件对比时差异更小, 甚至消失。虽然本研究在实验条件的设置上与前人研究类似, 但时频分析是对不同时间点, 不同频率区间能量大小的计算, 这一指标直接反映了不同活动速率的神经元激活/抑制情况(Cohen, 2017)。对于本研究而言, 我们没有发现 3~30 Hz 区间内其他频段的能量活动在音节相关条件与音素相关条件之间存在显著差异, 表明音节相关和音素相关两种条件之间存在的重叠度差异并没有混淆实验的主要结果。而且, 相比音节无关条件, 个体在音节相关条件下 θ 频段的活动能量更低; 但对音素而言, 相关条件下 θ 频段的活动能量有升高的趋势。这种实验条件之间相反的神经振荡模式反映的是个体在口语产生过程中对音节和音素的加工存在不同的认知神经机制。

综上, 我们认为汉语口语产生中 θ 频段的活动反映了对音节的加工, 从频段能量变化的角度为音节是汉语口语词汇产生中音韵编码的单元提供了支持证据。我们的研究也表明人类的语言理解和语言产生过程在神经振荡上引发了相同的能量活动变化。未来还需结合不同的实验范式, 进一步探索汉语口语产生中各个频段能量活动的认知意义。

参 考 文 献

Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102*(3), 464–475.

Bastiaansen, M., Magyari, L., & Hagoort, P. (2009). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience, 22*(7), 1333–1347.

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language, 141*, 62–69.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*(5), 512–528.

Brookes, M. J., Gibson, A. M., Hall, S. D., Furlong, P. L., Barnes, G. R., Hillebrand, A., ... Morris, P. G. (2005). GLM-beamformer method demonstrates stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. *NeuroImage, 26*(1), 302–308.

Burki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language, 150*, 90–102.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE, 5*(6), e10729.

Cai, X., Yin, Y. L., & Zhang, Q. F. (2020). The roles of syllables and phonemes during phonological encoding in Chinese spoken word production: A topographic ERP study. *Neuropsychologia, 144*, 1–10.

Chen, J. -Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient, 43*(1), 15–26.

Chen, J. -Y., Chen, T. -M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781.

Chen, J. -Y., O'Seaghdha, P. G., & Chen, T. -M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*(5), 825–836.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, M. X. (2017). Where Does EEG Come From and What Does It Mean. *Trends in Neurosciences, 40*(4), 208–218.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*(1), 119–132.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57*(2), 195–209.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*(2), 345–361.

Dell'acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology, 1*, 177.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D.

(2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*, 761–768.

Engell, A. D., & Mccarthy, G. (2014). Repetition suppression of face-selective evoked and induced EEG recorded from human cortex. *Human Brain Mapping, 35*(8), 4155–4162.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience, 12*(2), 105–118.

Feng, C., Yue, Y., & Zhang, Q, F. (2019). Syllables are Retrieved before Segments in the Spoken Production of Mandarin Chinese: An ERP Study. *Scientific Reports, 9*(1), 1–9.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*(1), 1–25.

Ghinst, M. V., Bourguignon, M., de Beeck, M. O., Wens, V., Marty, B., Hassid, S., ... de Tiege, X. (2016). Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *The Journal of Neuroscience, 36*(5), 1596–1606.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Goupillaud, P. L., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration, 23*(1), 85–102.

Grillspector, K., Henson, R. N., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14–23.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P. G., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biology, 11*(12), e1001752.

Gruber, T., & Müller, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Cognitive Brain Research, 13*(3), 377–392.

Gruber, T., Giabbiconi, C. M., Trujillobarreto, N. J., & Müller, M. M. (2006). Repetition suppression of induced gamma band responses is eliminated by task switching. *European Journal of Neuroscience, 24*(9), 2654–2660.

Howard, M. F., & Poeppel, D. (2012). The neuromagnetic response to spoken sentences: Co-modulation of theta band amplitude and phase. *Neuroimage, 60*(4), 2118–2127.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144.

Jacobs, C. L., & Dell, G. S. (2014). 'Hotdog', not 'Hot' 'dog': The phonological planning of compound words. *Language, Cognition and Neuroscience, 29*(4), 512–523.

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences, 30*(7), 317–324.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences, 16*(12), 606–617.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language, 148*, 51–63.

Li, X., Shao, X., Xia, J., & Xu, X. (2019). The cognitive and neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension. *Journal of Neurolinguistics, 49*, 155–167.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010.

Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Neural Information Processing Systems, 8*(8), 145–151.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*(1), 69–89.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping, 37*(8), 2767–2783.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence & Neuroscience, 2011*, 156869.

O'Seaghdha, P. G., Chen, J.- Y., & Chen, T.- M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282–302.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A. L. (2017). θ-band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience, 37*(33), 7930–7938.

Peña, M., & Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience, 24*(5), 1149–1164.

Perrin, F., Pernier, J., Bertrand, O. F., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*(2), 184–187.

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology, 30*(6), 547–558.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245–255.

Power, A. J., Mead, M., Barnes, L., & Goswami, U. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology, 3*, 216.

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the United States of America, 109*(35), 14265–14270.

Roelofs, A. (1997). The weaver model of word-form encoding

in speech production. *Cognition, 64*(3), 249–284.

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*(1), 22–37.

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences, 18*(1), 16–25.

Rugg, M. D., & Coles, M. G. H. (1995). *Oxford psychology series, No. 25. Electrophysiology of mind: Event-related brain potentials and cognition.* New York, NY, US: Oxford University Press.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Kupper, H., Kellmeyer, P., ... Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage, 49*(4), 3187–3197.

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition, 106*(2), 952–962.

Sereno, J. A., & Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. *Language and Speech, 58*(2), 131–151.

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience, 13*(2), 121–134.

van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190.

Wang, J., Wong, W. K., Wang, S., & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports, 7*(1), 5815.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences, 7*(12), 553–559.

You, W. P., Zhang Q. F., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS ONE, 7*(10), e46595.

Yu, M., Mo, C. & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS One, 9*(9), e106486.

Yue, Y., & Zhang, Q. F. (2015). Syllable and Segments Effects in Mandarin Chinese Spoken Word Production. *Acta Psychologica Sinica, 47*(3), 319–328.

[岳源, 张清芳. (2015). 汉语口语产生中音节和音段的促进和抑制效应. *心理学报, 47*(3), 319–328.]

Zhang, Q. F. (2005). The syllable's role in language production. *Advances in Psychological Science, 13*(6), 752–759.

[张清芳. (2005). 音节在语言产生中的作用. *心理科学进展, 13*(6), 752–759.]

Zhang, Q. F., & Damian, M. F. (2019). Syllables constitute proximate units for Mandarin speakers: Electrophysiological evidence from a masked priming task. *Psychophysiology, 56*(4), e13317.

Zhang, Q. F., & Wang, C. (2014). Syllable frequency and word frequency effects in spoken and written word production in a non-alphabetic script. *Frontiers in Psychology, 5*, 120.

Zhang, Q. F., & Yang, Y. F. (2003a). The determiners of picturenaming latency. *Acta Psychologica Sinica, 35*(4), 447–454.

[张清芳, 杨玉芳. (2003a). 影响图画命名时间的因素. *心理学报, 35*(4), 447–454.]

Zhang, Q. F., & Yang, Y. F. (2003b). The lexical access theory in speech production. *Journal of Developments in Psychology, 11*(1), 6–11.

[张清芳, 杨玉芳. (2003b). 言语产生中的词汇通达理论. *心理科学进展, 11*(1), 6–11.]

Zhang, Q. F., & Yang, Y. F. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science, 28*(2), 374–378.

[张清芳, 杨玉芳. (2005). 汉语单音节词汇产生中音韵编码的单元. *心理科学, 28*(2), 374–378.]

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language, 144*, 16–25.

Theta band (4~8 Hz) oscillations reflect syllables processing in Chinese spoken word production

JIANG Yuchen; CAI Xiao; ZHANG Qingfang

(*Department of Psychology, Renmin University of China, Beijing* 100872, *China*)

Abstract

Languages may differ in the proximate units of phonological encoding in spoken word production. It has been demonstrated that syllables are proximate units of phonological encoding in Chinese speech production. Previous studies report that the θ band oscillations has been associated with syllables processing in language comprehension, however, it remains unknown what are the neural oscillations for syllables retrieval in speech production. The present study aims to investigate the neural oscillations of syllables retrieval at the stage of phonological encoding in Chinese spoken word production.

We employed a masked priming paradigm and electrophysiological signals were recorded concurrently. In the task, participants were instructed to name pictures with disyllabic words which were preceded by briefly presented and masked prime words. Prime words were syllabically or phonemically related to the first syllable of targets or were unrelated, and the experimental design includes prime type (syllable vs. phoneme), relatedness

(related vs. unrelated), and repetition (first vs. second).

Behavioral data analysis showed a significant triple interaction among prime type, relatedness and repetition. In the first repetition, naming latencies were faster in syllabically related than unrelated condition, whereas latencies were longer in the phonemically related than unrelated condition. Time-frequency analysis also showed a significant interaction between prime type and relatedness in the time window of 300-600 ms after pictures onset. Specifically, theta band power was lower for syllabically related than unrelated while no significant differences between phonemically related and unrelated. Cluster based permutation test showed that, in the first repetition, syllabically related condition elicited lower θ band power than unrelated in the time window of 270-460 ms, while phoneme relatedness produced marginally higher θ band power than phoneme unrelatedness in the 340-390 ms time window. These effects were absent in the second repetition.

In sum, we found that syllable priming effect was reflected by the decrease of θ oscillation in spoken word production in Chinese. Time-frequency analysis also revealed an early syllable priming effect, and a late phonemic inhibition effect in Chinese spoken word production, which provides evidence for the proximate units principle.

Key words speech production; syllable; theta band; time-frequency analysis